Australian Government

ANSTO

The next generation of Scientific Computing @AS

AFAD 2023

Dr Andreas Moll

Manager – Scientific Computing

Science. Ingenuity. Sustainability.

Who we are

Scientific Computing founded in June 2017

Support Science and Users

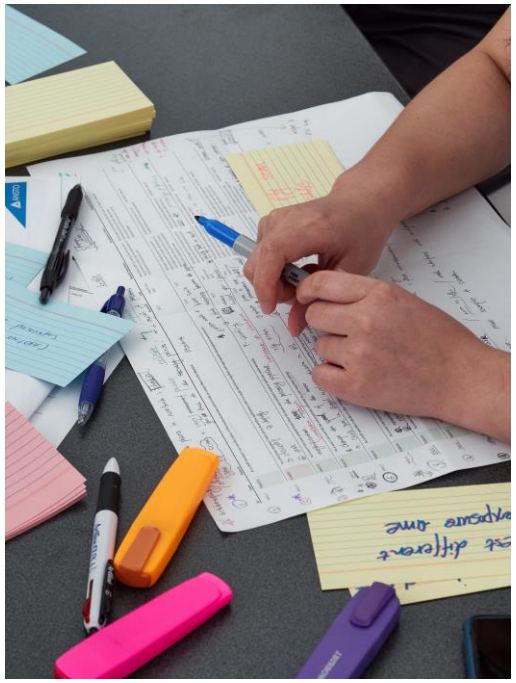
- Experiment Control
- Data Acquisition
- Data Processing
- Data Analysis

Our Team

- 1 manager
- 20 members
  - 1 principal engineer
  - 2 DevOps
  - 10 PhDs
  - 50% gender split


Australian Synchrotron
Clayton, Victoria


User Stories



Australian Synchrotron & BRIGHT

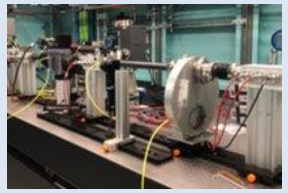
A user focused facility

- 5500+ visits per year
- 10 (+2) operating beamlines
- 586 Journal Publications in 2022
- Generate 1.5 PB of data each year

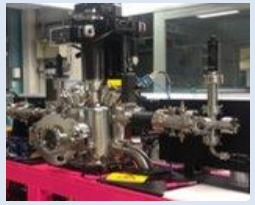
A growing facility (BRIGHT program)

- 8 new beamlines
- 2 already operating
- Visits and data size going to double
- Opportunity to “refresh” software

Micro-CT
(8 – 40 keV)
Sept 2022



MEX-1
(3.5 - 13.6 keV)
Nov 2022

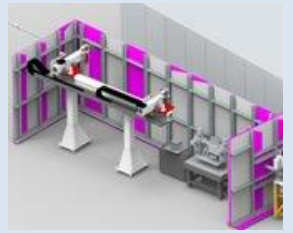


MEX-2
(1.3 - 3.5 keV)
Mid-2023

BioSAXS
(8 – 15 keV)
Mid-2023

MX3
(10-15 keV)
Mid-2024

ADS-1
(50 – 150 keV)
Late-2024

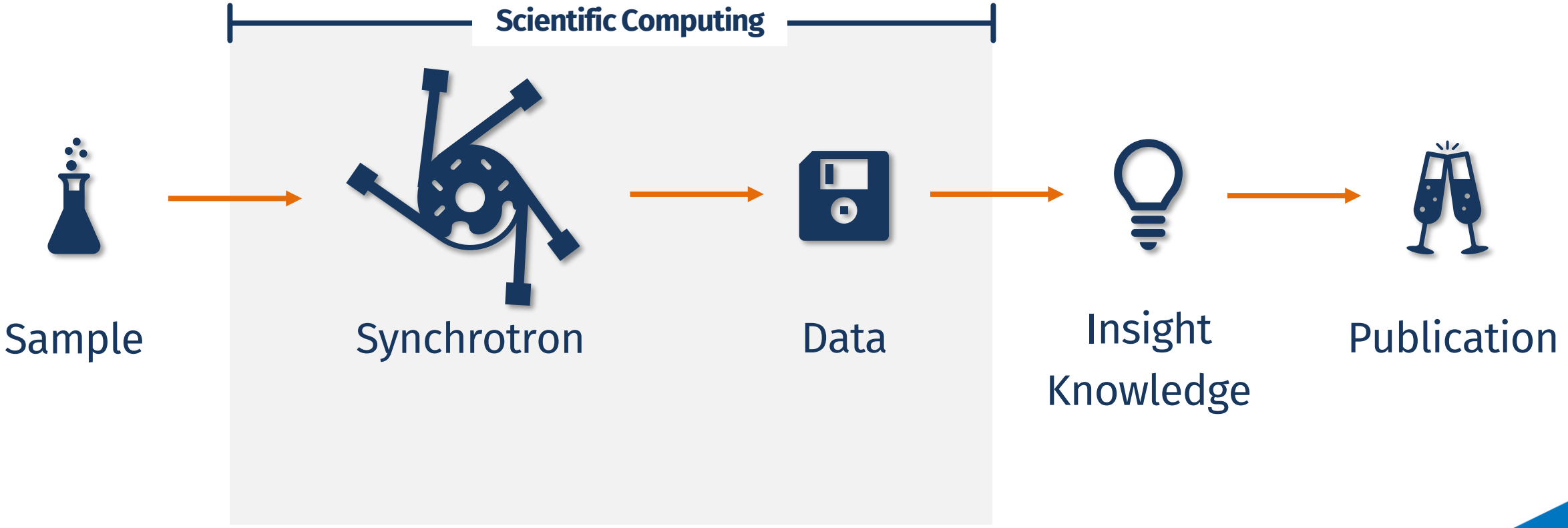


ADS-2
(45 – 90 keV)
Late-2024

NANO
(5 – 18 keV)
Mid-2025

ANSTO

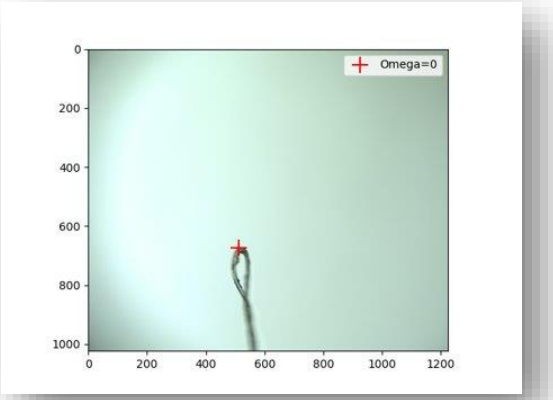
A typical User Journey



Samples & Robots

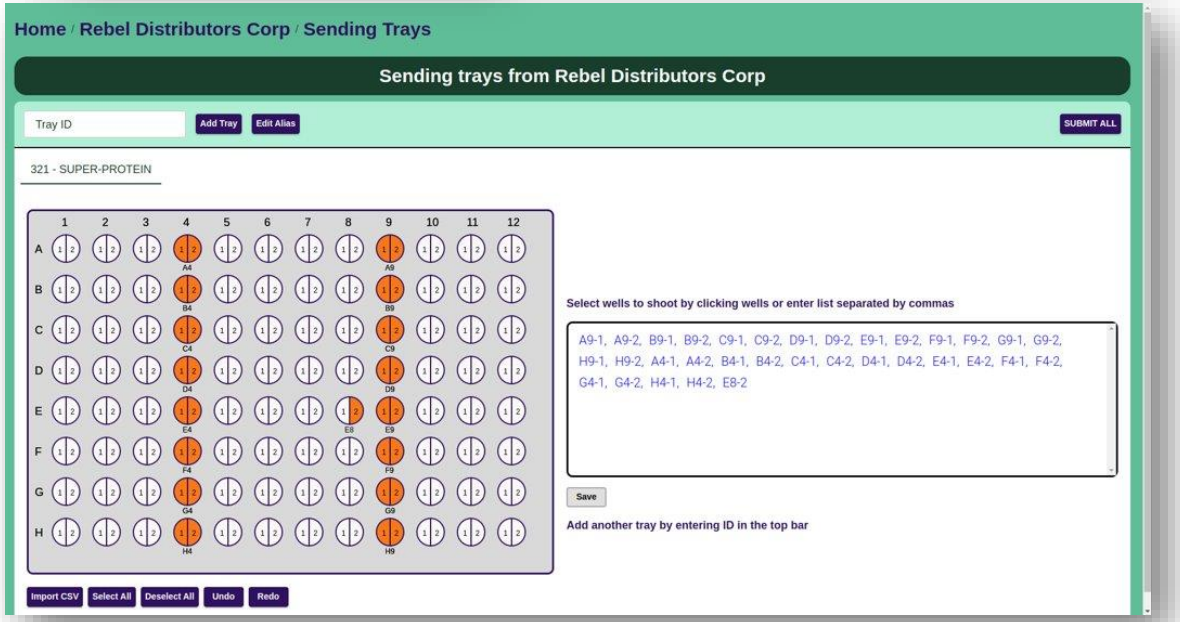
High throughput

- Robots loading samples
- Large number of samples
- Samples organised in trays, pucks, ...
- Connect samples to processing results



User friendly and fast

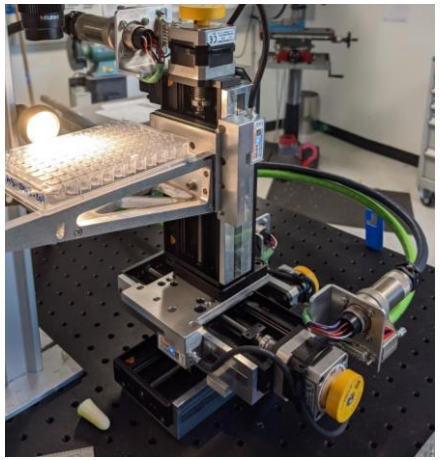
- Modern, web-based sample management applications
- Efficient sample changes
- Automated sample screening and quality checks



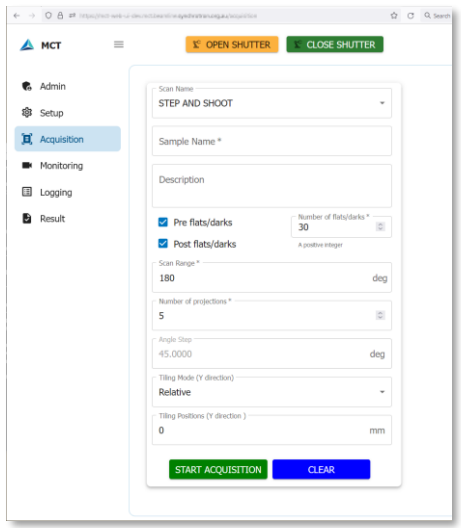
python™ FastAPI PostgreSQL

Data Collection

Hardware



python™ bluesky



Data Collection

FastAPI

kafka

ØMQ

redis

Local Storage

HDF

Flexible experiment control

- Ophyd and Bluesky
  - Python abstraction layer and orchestration
  - developed by NSLS-II
  - sits on top of EPICS
- Controlled via
  - Web GUI for “normal” users
  - Jupyter notebook for “expert” user

High throughput

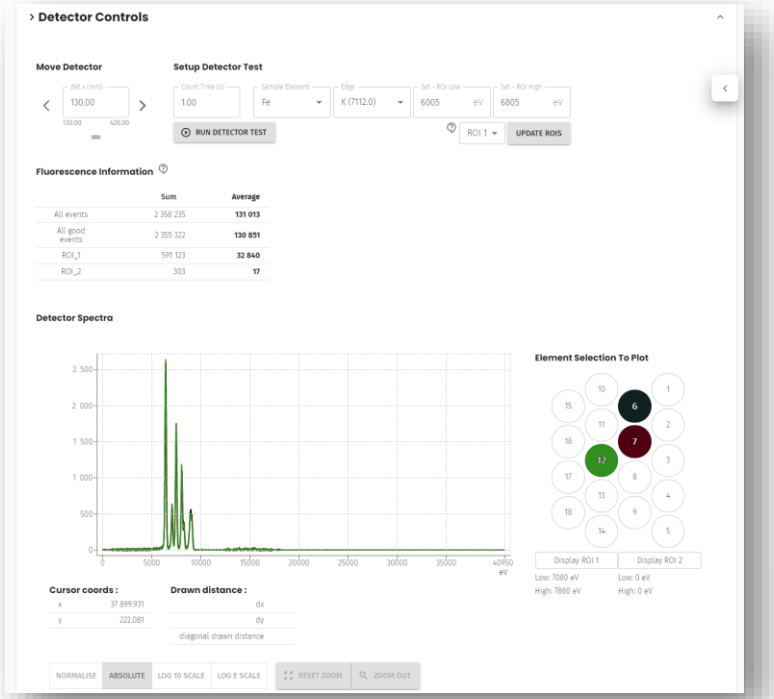
- Stream data (avoid disk if possible)
- Capture rich metadata
- Store as “data product”

ANSTO

Web GUI

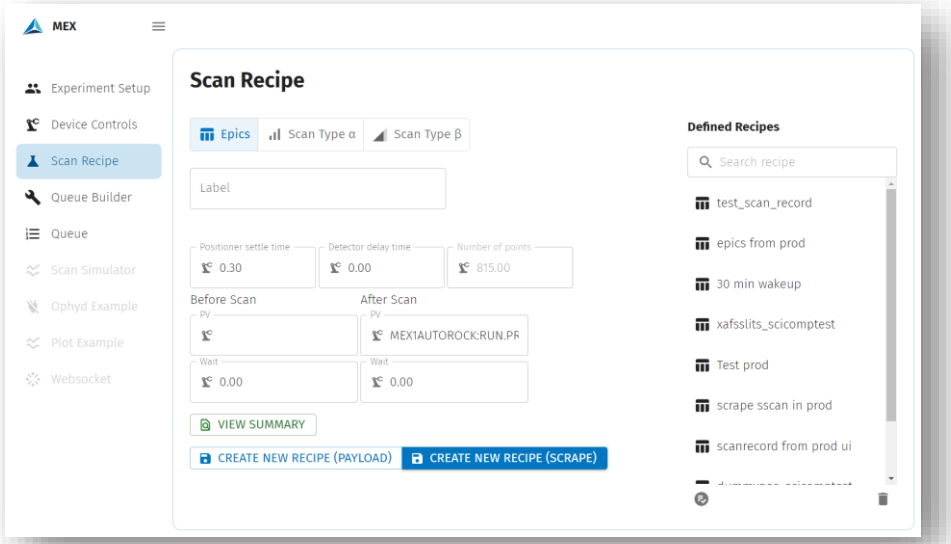
Modern interfaces

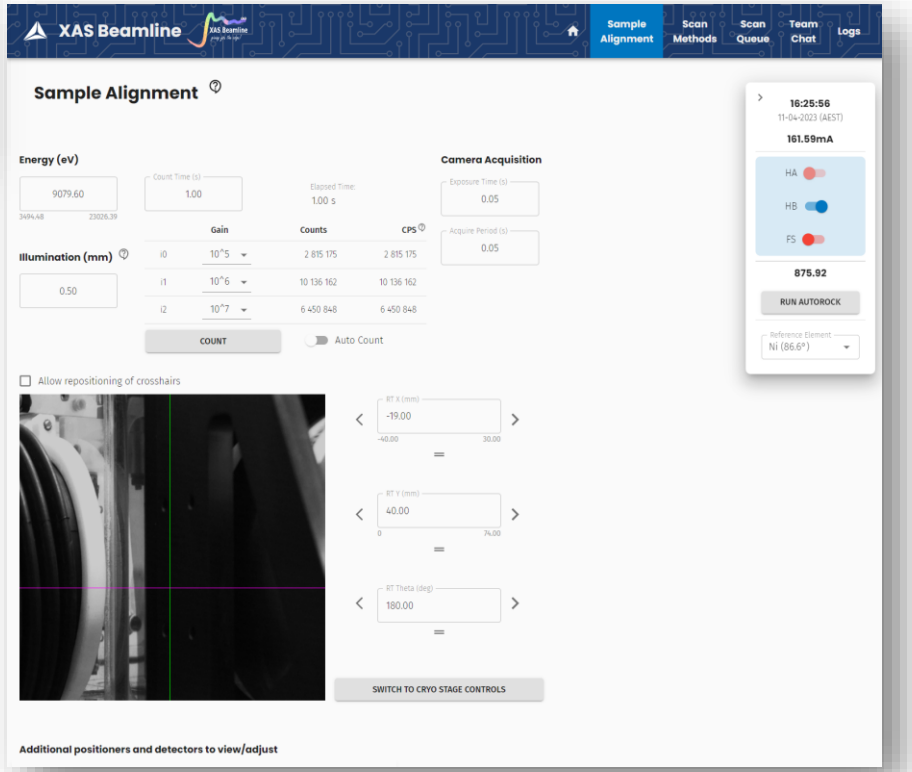
- Web-based GUIs
  - No local installation required
  - Remote access ready
- Established design language
  - Material UI
  - Reduces training time for users



React

TypeScript



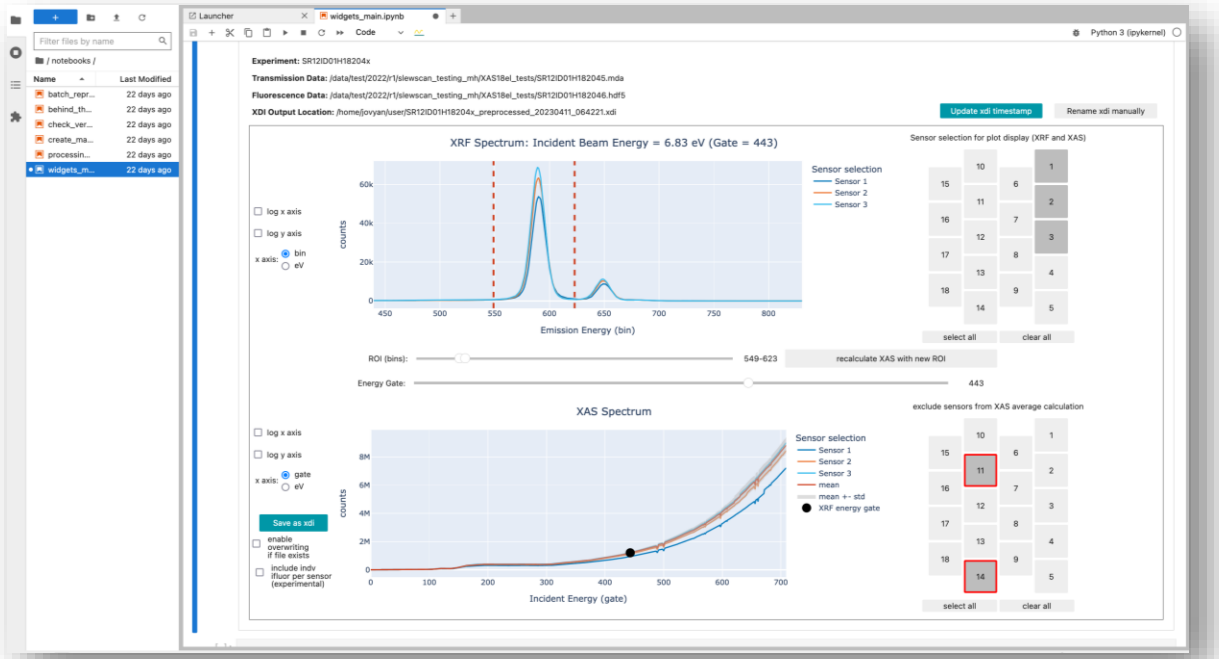


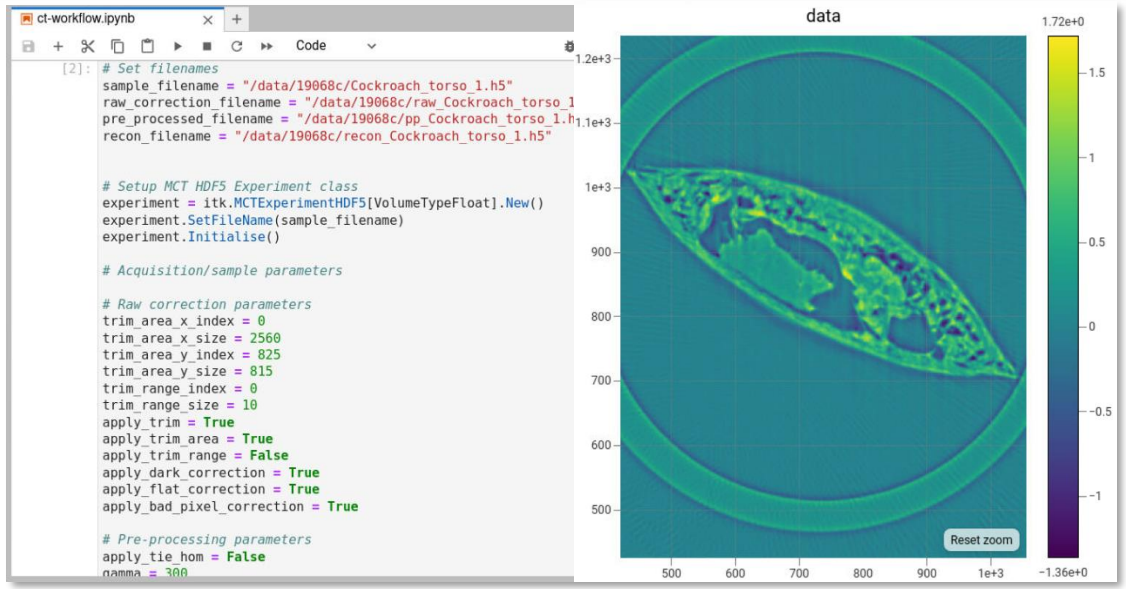
Data Processing

High performance computing

- HPC system
  - Comprised of 50+ physical servers
  - GPU nodes for image processing

- Beamline specific implementation
  - Heavily depends on beamline
  - Off-the-shelf and in-house tools
  - Heavy use of frameworks (e.g. ITK, Prefect)
  - Runs in Docker containers on Kubernetes

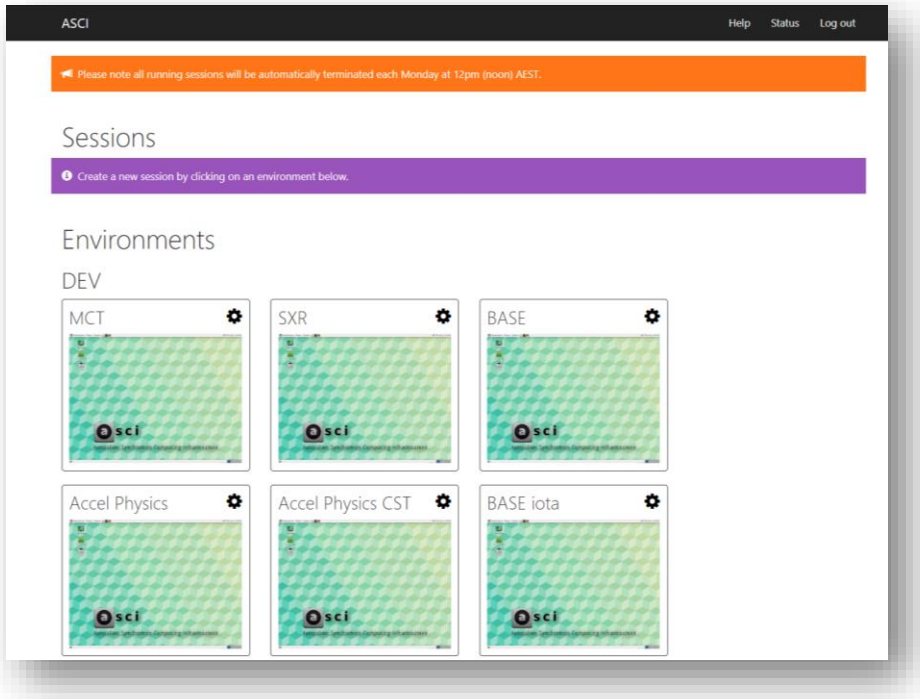




Data Analysis & Visualisation

Remote analysis platform

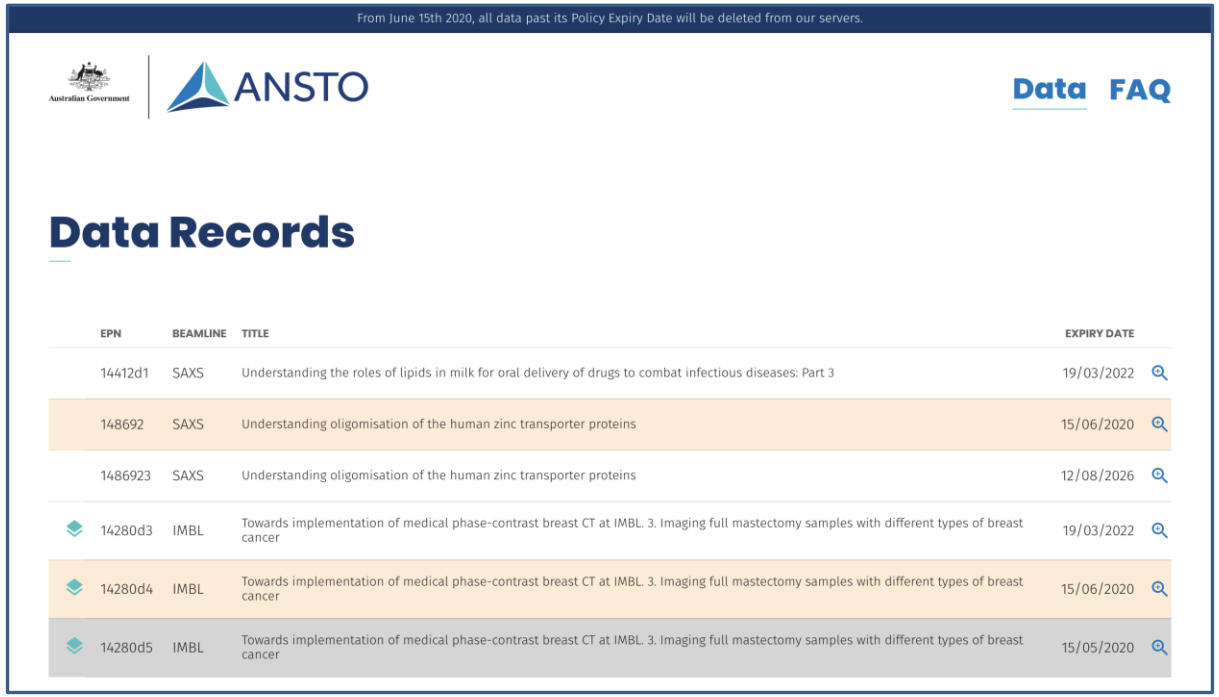
- Remote desktop environment in browser
- User starts session with tools pre-installed and data mounted



Data Management

We are only the temporary custodians of data

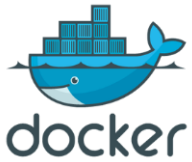
- Data has a lifecycle
  - Generated data is organised by datasets of “data products”
  - Our standard data container is HDF5
  - We delete data after 1 or 3 years (depends on beamline)
  - Space for 6PB locally
- Data transport
  - Users need to take their data home
  - We offer SFTP service
  - We will offer Globus in the future



HDF

ANSTO

Infrastructure



Private cloud

- DevOps & Containerisation
  - Everything runs in a Docker container
  - Fully automated CI/CD pipeline
  - Dedicated DevOps engineer
  - Infrastructure as Code
  - No ssh'ing into VMs, no creation of snowflake machines

- Industry standard orchestration
  - All containers run in Kubernetes (including EPICS clients)
  - Deployed through Helm charts
  - Logging through ELK
    - Currently investigating Sentry